

## **IV Ogólnopolska Konferencja EBIB Internet w bibliotekach Open Access**

**Toruń, 7-8 grudnia 2007 roku**

**Agnieszka Lewandowska, Cezary Mazurek, Marcin Werla  
Poznańskie Centrum Superkomputerowo-Sieciowe**

### **Federacja Bibliotek Cyfrowych w sieci PIONIER - Dostęp do otwartych bibliotek cyfrowych i repozytoriów**

#### ***Federation of Digital Libraries working in the PIONIER network – an access to open digital libraries and repositories***

##### **Abstrakt**

Badania w zakresie bibliotek cyfrowych rozpoczęto w Poznańskim Centrum Superkomputerowo-Sieciowym już w 1996 roku. W ramach tych prac od 1999 roku rozwijane jest oprogramowanie dLibra – najpopularniejsze obecnie oprogramowanie do budowy bibliotek cyfrowych w Polsce. Obecnie w sieci PIONIER uruchomionych jest łącznie 16 bibliotek cyfrowych opartych na tym oprogramowaniu, w tym największa w Polsce Wielkopolska Biblioteka Cyfrowa. Biblioteki te razem dają dostęp do prawie 80 000 zróżnicowanych obiektów cyfrowych. Tak duża liczba bibliotek i zasobów cyfrowych umożliwiła budowę na ich bazie zaawansowanych usług sieciowych. Pierwszą z takich usług było wyszukiwanie rozproszone dostępne w oprogramowaniu dLibra od lutego 2006 roku. W czerwcu bieżącego roku udostępniony został serwis Federacja Bibliotek Cyfrowych (FBC, <http://fbc.pionier.net.pl/>). Jego głównym celem jest ułatwienie korzystania z zasobów cyfrowych dostępnych w sieci PIONIER i zwiększenie ich widoczności na świecie. Serwis ten obecnie daje swoim użytkownikom możliwość skorzystania z trzech usług:

1. wyszukiwania rozproszonego w dostępnych publikacjach,
2. wyszukiwania rozproszonego w planach digitalizacji,
3. tworzenia i rozpoznawania trwałych referencji do obiektów cyfrowych opartych o ich unikalne identyfikatory OAI.

Serwis FBC oparty jest na otwartych protokołach komunikacyjnych (OAI-PMH, RSS, OpenSearch) i jest w stanie współpracować z dowolnymi bibliotekami oraz repozytoriami cyfrowymi, które te protokoły wspierają. Taka współpraca nie wymaga przekazywania zasobów cyfrowych na rzecz FBC – pozostają one cały czas w macierzystych repozytoriach. W niniejszym artykule chcielibyśmy opisać możliwości, jakie serwis FBC daje użytkownikom polskich zasobów cyfrowych oraz twórcom bibliotek i repozytoriów, w których zasoby te są gromadzone, a także przedstawić plany dalszego rozwoju FBC.

##### **Abstract**

*Poznań Supercomputing and Networking Center started its research on digital libraries in 1996. Within that project since 1999 the dLibra software has been developed – at present the most popular digital libraries software. 16 digital libraries built on aforementioned software function in the PIONIER network, including the biggest one – Digital Library of Wielkopolska. The libraries altogether offer access to about 80.000 digital objects of different types. Thanks to such a huge number of libraries and digital resources it was possible to create advanced network services. The first service was a distributed searching available in dLibra software since 2006. In June 2007 the “Federation of Digital Libraries” service was launched (FBC, <http://fbc.pionier.net.pl/>). Its main goal is to facilitate the usage of digital resources available in the Pionier network and make them more visible in the world. At present the service offers three options to its users:*

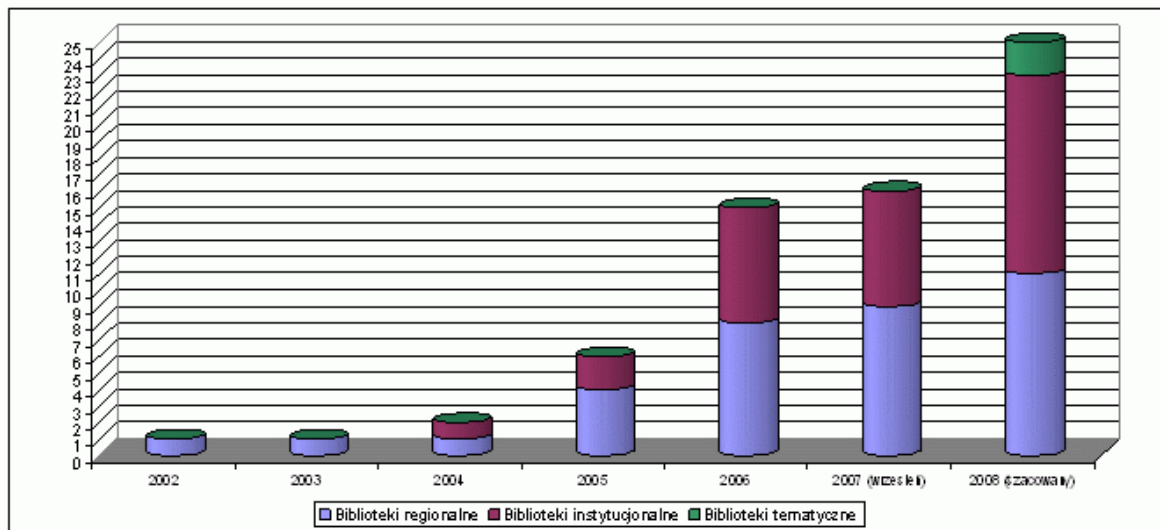
1. *Distributed searching in available publications*

2. *Distributed searching in digitization plans*
3. *Creation and recognition of permanent references to digital object using their unique OAI identifiers.*

*FBC service is based on open communication protocols (OAI-PMH, RSS, Open Search) and it is able to co-operate with each type of digital library and repository supported by those protocols. Such a co-operation does not need digital resources transfer to FBC – they stay in their home repositories all the time. In this paper we would like to describe the options FBC service offers to the Polish users of digital resources as well as the creators of libraries and repositories in which the digital content is collected. We also would like to present the plans of FBC development.*

## Wprowadzenie

W 2002 roku oficjalnie udostępniono w Internecie Wielkopolską Bibliotekę Cyfrową (WBC) – pierwszą tego typu inicjatywę w Polsce. Podstawowym celem WBC, współtworzonej przez poznańskie środowisko akademickie oraz biblioteki naukowe i publiczne Poznania i Wielkopolski, było udostępnianie przez Internet cyfrowych publikacji zgromadzonych w czterech kolekcjach tematycznych: materiały dydaktyczne, dziedzictwo kulturowe, materiały regionalne oraz muzykalia. Koordynacją działań związanych z WBC zajęła się Poznańska Fundacja Bibliotek Naukowych, a za stronę techniczną odpowiedzialne zostało Poznańskie Centrum Superkomputerowo-Sieciowe (PCSS). Instalację WBC oparto na oprogramowaniu dLibra, dzięki czemu możliwe było równoczesne tworzenie biblioteki cyfrowej w sposób rozproszony (tj. przez wiele różnych instytucji, w oparciu o wymianę danych poprzez Internet), spójne i efektywne przechowywanie i udostępnianie gromadzonych zasobów cyfrowych, a także elastyczne zarządzanie nimi<sup>[1]</sup>. Obecnie WBC zawiera około 43 000 cyfrowych publikacji, co czyni ją największą biblioteką cyfrową w Polsce. Przy obecnym tempie udostępniania nowych zasobów (około 1 500 – 2 000 obiektów miesięcznie) do końca 2007 roku WBC powinna przekroczyć próg 50 000 publikacji.



Rys. 1. Liczba publicznie dostępnych bibliotek cyfrowych opartych na oprogramowaniu dLibra (o ile nie zaznaczono inaczej, dane na koniec poszczególnych lat).

Uruchomienie WBC zapoczątkowało proces tworzenia kolejnych bibliotek cyfrowych. W 2004 roku powstała Biblioteka Cyfrowa Politechniki Wrocławskiej, przekształcona później w Dolnośląską Bibliotekę Cyfrową. Poza WBC, największe obecnie biblioteki cyfrowe to: Kujawsko-Pomorska Biblioteka Cyfrowa (ok. 15 000 publikacji), Biblioteka Cyfrowa Uniwersytetu Wrocławskiego (ok. 10 000 publikacji) oraz Śląska Biblioteka Cyfrowa, Małopolska Biblioteka Cyfrowa i Cyfrowa Biblioteka Narodowa POLONA (każda po ok. 5 000 publikacji). Łącznie wszystkie biblioteki cyfrowe udostępniają około 90 000 cyfrowych dokumentów. Dynamika wzrostu liczby bibliotek cyfrowych w Polsce

przedstawiona jest na rysunku 1. Obecnie funkcjonuje 16 bibliotek cyfrowych opartych na oprogramowaniu dLibra, spośród których 9 to regionalne biblioteki cyfrowe, a 7 to biblioteki instytucjonalne. W roku 2008 prawdopodobnie udostępnione zostaną pierwsze tematyczne biblioteki cyfrowe. Treści do tych bibliotek opracowywane są przez ponad 150 instytucji.

Tak duża liczba bibliotek cyfrowych oraz treści cyfrowych dostępnych on-line stały się podstawą do rozpoczęcia prac nad stworzeniem jednolitej infrastruktury działającej w skali całego kraju. W dążeniach nad ujednoczeniem infrastruktury pomógł projekt badawczy nr 3 T11C 023 30 Ministerstwa Nauki i Szkolnictwa Wyższego: „Mechanizmy usług atomowych dla rozproszonych bibliotek cyfrowych” zrealizowany przez PCSS. Dzięki pracom badawczo-rozwojowym uruchomione w sieci Polski Internet Optyczny PIONIER biblioteki cyfrowe zostały połączone przy pomocy protokołu OAI-PMH <http://www.openarchives.org/OAI/openarchivesprotocol.html>. Służący do wymiany metadanych protokół umożliwił zrealizowanie mechanizmu wyszukiwania rozproszonego, pozwalającego na przeszukiwanie opisów zasobów cyfrowych wszystkich bibliotek opartych na oprogramowaniu dLibra z poziomu ich stron WWW[2]. Wykorzystanie protokołu OAI-PMH spowodowało również powstanie automatycznie nadawanych, unikalnych w skali światowej, identyfikatorów udostępnianych obiektów cyfrowych.

Kolejny element infrastruktury bibliotek cyfrowych w Polsce to wyszukiwarka Federacja Bibliotek Cyfrowych (FBC) udostępniona w czerwcu 2007 roku pod adresem <http://fbc.pionier.net.pl/>. Poniżej opisano ideę tego serwisu, zasadę jego działania oraz podstawowe funkcje. Na końcu artykułu przedstawiono również trwające obecnie prace nad rozwojem FBC i sieci polskich bibliotek cyfrowych.

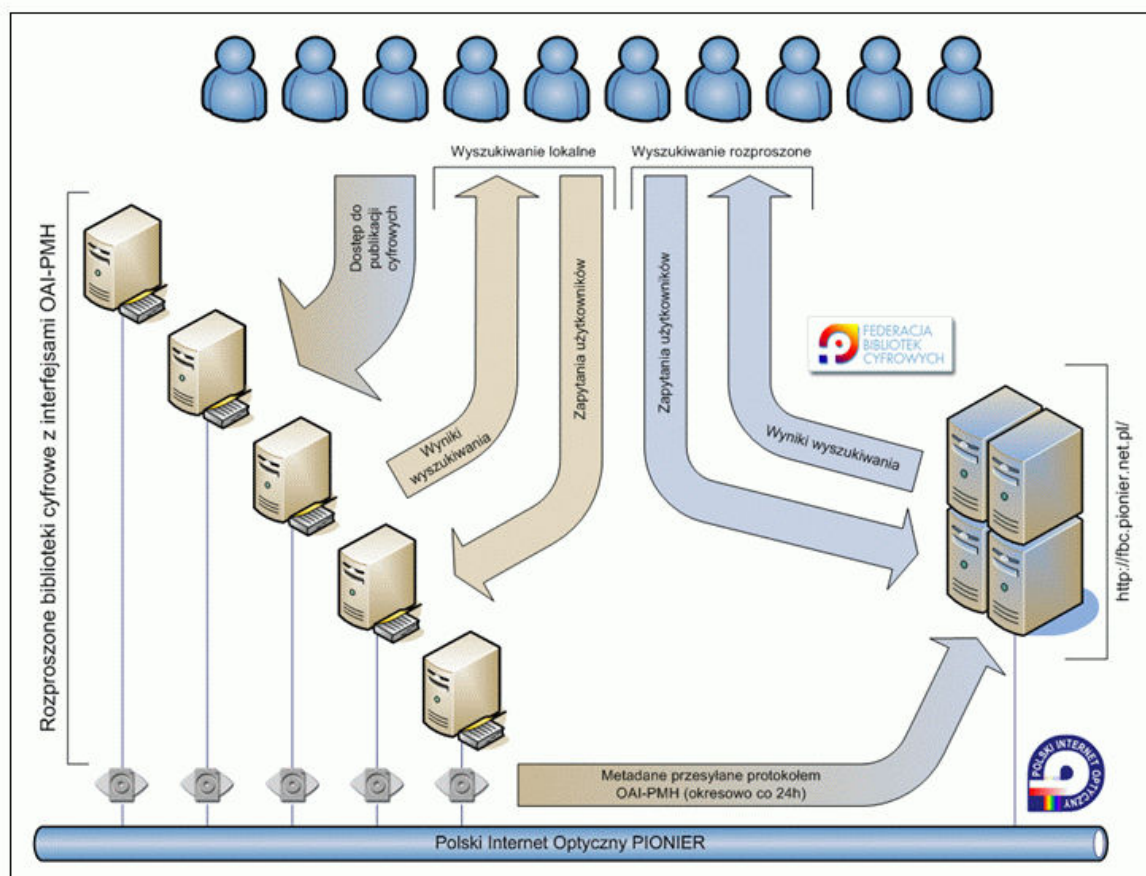
## **Serwis Federacja Bibliotek Cyfrowych**

### **Informacje ogólne**

Federacja Bibliotek Cyfrowych ma na celu wirtualne połączenie bibliotek cyfrowych i repozytoriów dostępnych w polskim Internecie oraz udostępnienie nowych zaawansowanych funkcji i usług sieciowych realizowanych w tym środowisku. Misją serwisu jest ułatwienie wykorzystania zasobów polskich bibliotek cyfrowych i repozytoriów oraz zwiększenie widoczności tych zasobów w światowym Internecie.

Serwis FBC jest utrzymywany i rozwijany przez PCSS. Jednym z podstawowych założeń technicznej koncepcji serwisu jest oparcie komunikacji z bibliotekami cyfrowymi i repozytoriami na otwartych protokołach i formatach danych (obecnie podstawą jest wspomniany wcześniej protokół OAI-PMH oraz schemat Dublin Core). Dzięki temu FBC nie jest związane z żadnym konkretnym dostawcą oprogramowania dla bibliotek i repozytoriów cyfrowych. Jedynym wymaganiem stawianym repozytorium, które mają być widoczne w FBC, jest możliwość komunikacji przy pomocy protokołu OAI-PMH.

Z formalno-organizacyjnego punktu widzenia współpraca bibliotek cyfrowych i repozytoriów z serwisem FBC jest bardzo prosta. Nie wymaga żadnych opłat ani dodatkowego nakładu pracy ze strony administratorów repozytorium, jak również przekazywania publikacji cyfrowych na rzecz FBC. Trzeba jedynie podać administratorom FBC adres URL pozwalający na pobieranie danych z repozytorium poprzez protokół OAI-PMH – czyli adres interfejsu OAI-PMH tego repozytorium. Pod tym względem serwis FBC działa tak samo jak inne bazujące na repozytoriach OAI-PMH serwisy np. OAlster <http://www.oaister.org/>. W ramach współpracy FBC okresowo, automatycznie pobiera metadane udostępnianych obiektów cyfrowych i gromadzi je w lokalnym katalogu. Pobrane z rozproszonych repozytoriów metadane stanowią podstawę działania serwisu FBC. Użytkownicy mogą przeszukiwać lokalne biblioteki cyfrowe, a równocześnie poprzez FBC mają dostęp do wyszukiwania rozproszonego (rysunek 2).



Rys. 2. Schemat obrazujący zasadę działania serwisu Federacja Bibliotek Cyfrowych.

## Wyszukiwanie rozproszone

Aby przeszukać zasoby bibliotek cyfrowych użytkownicy mają teraz dwie możliwości: mogą użyć wyszukiwania lokalnego po kolei w każdej z bibliotek cyfrowych i analizować uzyskane w ten sposób listy wyników lub skorzystać z wyszukiwania rozproszonego w FBC albo w innej bibliotece cyfrowej. Stosując wyszukiwanie rozproszone otrzymują jedną listę wyników zawierającą zasoby z wielu repozytoriów. Wybranie konkretnego wyniku z tej listy powoduje przejście do właściwej strony biblioteki czy repozytorium, w którym ten obiekt się znajduje. Przykładowe wyniki wyszukiwania przedstawiono na rysunku 3.

Dodatkowo, w celu jak najszerszego promowania polskich zasobów cyfrowych, serwis FBC pozwala na umieszczenie swojego interfejsu wyszukiwawczego na każdej dowolnej stronie WWW w postaci mini-wyszukiwarki. Wtedy internauci odwiedzający tę stronę mogą bezpośrednio z niej przeszukać zasoby polskich bibliotek cyfrowych (rysunek 4). Osadzenie na własnej stronie mini-wyszukiwarki jest bardzo łatwe i sprowadza się do nieznacznego zmodyfikowania kodu HTML tej strony. Ponadto każdy internauta może również dodać funkcję przeszukiwania FBC bezpośrednio do swojej przeglądarki internetowej. Rozszerzenie to jest oparte na formacie OpenSearch <http://www.opensearch.org/Specifications/OpenSearch/1.1>, który obecnie jest obsługiwany przez przeglądarki Firefox 2 i Internet Explorer 7. Szczegółowe informacje dostępne są na stronach FBC w sekcji „Dodatki” <http://fbc.pionier.net.pl/owoc/text?id=addons>.





Rys. 3. Przykładowe wyniki wyszukiwania w serwisie FBC (<http://fbc.pionier.net.pl>, dostęp 15.09.2007 roku).



Rys. 4. Strona katalogów Biblioteki Kórnickiej PAN z osadzoną mini wyszukiwarką FBC (<http://www.bkpan.poznan.pl/katalogi/index.html>, dostęp 15.09.2007 roku).

## Trwałe referencje do obiektów cyfrowych

Jak wspomniano wcześniej, udostępnienie protokołu OAI-PMH w sieci polskich repozytoriów i bibliotek cyfrowych zaowocowało powstaniem automatycznie nadawanych, unikalnych w skali światowej identyfikatorów udostępnianych obiektów. Mechanizm tych identyfikatorów – OAI Identifier – zawarty jest w standardzie opisującym protokół OAI-PMH. Format identyfikatora wygląda następująco:

oai:<domena repozytorium>:<identyfikator zasobu w repozytorium>

a przykładowa jego instancja:

oai:www.wbc.poznan.pl:8711

Serwis FBC posiada mechanizm rozpoznawania identyfikatorów OAI Identifier, a tym samym pozwala na uzyskanie informacji oraz aktualnego adresu obiektu cyfrowego na podstawie jego unikalnego identyfikatora. Mechanizm ten może być również wykorzystany do utworzenia trwałej referencji do obiektu cyfrowego np. na potrzeby odwołań w bibliografii. Referencja taka ma następującą postać:

<http://fbc.pionier.net.pl/id/>

a przykładowe odwołanie wygląda w ten sposób:

<http://fbc.pionier.net.pl/id/oai:www.wbc.poznan.pl:8711>

Otwarcie zamieszczonego powyżej adresu w przeglądarce WWW spowoduje wyświetlenie podstawowych metadanych obiektu cyfrowego o podanym identyfikatorze oraz odnośników do metadanych i/lub treści tego obiektu. Utworzona w ten sposób referencja może być trwałym i poprawnym odwołaniem do cyfrowej publikacji, niezależnie od zmian wprowadzonych w oprogramowaniu biblioteki cyfrowej, która tę publikację udostępnia. Informacje o identyfikatorach poszczególnych obiektów cyfrowych, które można wykorzystać do tworzenia referencji zazwyczaj udostępniane są przez poszczególne repozytoria na stronach z opisem obiektów.

## Koordynacja digitalizacji

Szczególnie istotną funkcją FBC, z punktu widzenia twórców polskich zasobów cyfrowych, jest możliwość przeszukiwania zebranych w jednym miejscu planów digitalizacji poszczególnych bibliotek cyfrowych. Ta funkcja bazuje na pobieranych przez protokół OAI-PMH informacjach o planowanych publikacjach poszczególnych bibliotek cyfrowych. Poza przeszukiwaniem możliwe jest również przeglądanie pełnej listy publikacji planowanych oraz wygenerowanie specjalnej postaci tej listy przeznaczonej do wydruku.

W chwili obecnej system dLibra (wersja 3.0) jest jedynym znanym autorem artykułu oprogramowaniem, które (od wersji 1.6 wydanej w sierpniu 2005 roku) wspiera planowanie digitalizacji i równocześnie udostępnia gromadzone w ten sposób dane poprzez protokół OAI-PMH. Cykl życia publikacji elektronicznej w bibliotece cyfrowej opartej na systemie dLibra może rozpoczynać się właśnie od „publikacji planowanej”, która pozwala na przechowanie metadanych na temat jeszcze nieistniejącego obiektu cyfrowego<sup>[3]</sup>. Publikację taką można później automatycznie przekształcić w pełny obiekt cyfrowy dołączając do niej pliki składowe tego obiektu. Jak wspomniano wcześniej, serwis FBC jest technicznie gotowy do współpracy z innymi systemami do budowy bibliotek cyfrowych, gdy tylko zaczną one gromadzić stosowne informacje i udostępnią je poprzez protokół OAI-PMH.

Liczba publikacji planowanych, o których informacje dostępne są w systemie FBC, waha się obecnie w zakresie od 600 do 800. Wahania te są naturalnym następstwem działań poszczególnych bibliotek polegających na przekształcaniu obiektów planowanych w pełne obiekty cyfrowe oraz dodawaniu nowych obiektów planowanych.

## Dalsze prace

Przedstawiona powyżej funkcjonalność serwisu FBC jest ciągle rozwijana i ulepszana. Najbliższe plany rozwoju serwisu obejmują mechanizmy związane m.in.: ze statystykami obejmującymi dostępność obiektów cyfrowych, z uatrakcyjnieniem prezentacji zgromadzonych metadanych (np. poprzez wykorzystanie „chmur tagów”) a także dodatek do FBC umożliwiający integrację z popularnym serwisem iGoogle. Poza tym trwają prace nad bardziej znaczącymi modyfikacjami serwisu związanymi z dalszą rozbudową infrastruktury sieci bibliotek cyfrowych i repozytoriów w Polsce. Pierwsza z nich polegać będzie na przekształceniu obecnej bazy publikacji planowanych w społecznościową platformę do koordynacji digitalizacji. Docelowo platforma ta ma umożliwiać m.in.: gromadzenie informacji o planach digitalizacji również z tych bibliotek cyfrowych, w których oprogramowanie nie pozwala na zbieranie i współdzielenie informacji oraz możliwości konsultacji planów digitalizacji z czytelnikami poprzez głosowanie na poszczególne planowane obiekty czy proponowanie własnych publikacji. Dyskusja na temat ostatecznego kształtu tej platformy toczy się na forum Biblioteka 2.0 (w wątku dostępnym pod adresem <http://forum.biblioteka20.pl/viewtopic.php?p=896>). Drugą znaczącą modyfikacją serwisu ma być włączenie go do, budowanej w ramach sieci PIONIER, rozproszonej platformy uwierzytelniania i autoryzacji dla bibliotek cyfrowych. Platforma ta pozwoli przede wszystkim na wykorzystanie jednego profilu użytkownika w wielu bibliotekach cyfrowych (i w serwisie FBC). Dzięki temu czytelnik będzie mógł np. na swojej wirtualnej półce z ulubionymi książkami umieszczać obok siebie publikacje pochodzące z różnych bibliotek cyfrowych, a półkę tę będzie miał dostępną w każdej bibliotece cyfrowej, do której się zaloguje. Rozszerzenie obejmie też wykorzystanie wielu różnych zewnętrznych baz użytkowników (np. baz studentów wyższych uczelni) przy autoryzacji dostępu do zasobów cyfrowych (np. skryptów akademickich). Podobne systemy budowane są również w innych krajach na świecie np. system ATHENS w Wielkiej Brytanii <http://www.athensams.net/>. Zazwyczaj skupiają się one jednak przede wszystkim na kwestiach autoryzacji dostępu do określonych zasobów (np. licencjonowanych czasopism elektronicznych) i nie dają funkcji sieciowego profilu czytelnika.

## Zakończenie

Opisana w niniejszym artykule Federacja Bibliotek Cyfrowych jest najnowszym elementem infrastruktury bibliotek cyfrowych w Polsce. Serwis ten rozwijany i utrzymywany jest przez Poznańskie Centrum Superkomputerowo-Sieciowe. Dostępne w systemie funkcje i dodatki pozwalają na efektywne wyszukiwanie w polskich bibliotekach cyfrowych, na przeszukiwanie i przeglądanie planów digitalizacji tych bibliotek oraz na łatwe identyfikowanie i odwoływanie się do poszczególnych obiektów cyfrowych. Twórcy serwisu planują rozwijać go tak, aby był on nowoczesnym narzędziem przydatnym zarówno dla użytkowników zasobów polskich bibliotek cyfrowych jak i dla twórców tych zasobów. Kierunki rozwoju serwisu FBC oraz całej sieci bibliotek cyfrowych w Polsce, zgodne będą również z kierunkami podobnych prac realizowanych w innych projektach na świecie. Działania takie pozwolą na funkcjonowanie jednej spójnej przestrzeni informacyjnej, której polskie biblioteki cyfrowe będą ważnym i nieodzownym elementem.

## Bibliografia:

[1] MAZUREK, C. STROIŃSKI, M. WERLA, M. Wdrażanie regionalnych bibliotek cyfrowych w sieci PIONIER w oparciu o środowisko dLibra. In *Materiały konferencyjne z IV Krajowej Konferencji Naukowej INFOBAZY 2005. Bazy danych dla nauki* [on-line]. Gdańsk: Centrum Informatyczne TASK; PG, 2005. [Dostęp 9 listopada 2007] s. 58-63. Dostępny w World Wide Web: <http://dlibra.psnc.pl/biblioteka/publication/16>. ISBN 83-908112-3-5.

[2] MAZUREK C. STROIŃSKI M. WERLA M. WĘGLARZ, J. Metadata harvesting in regional digital libraries in PIONIER Network. *Campus-Wide Information Systems* 2006, Vol. 23, no. 4, s. 241-253. ISSN 1065-0741.

[3] MAZUREK, C. WERLA, M. Digital Object Lifecycle in dLibra Digital Library Framework. In *Proceedings of the 9th International Workshop of the DELOS Network of Excellence on Digital Libraries on Digital Repositories, Heraklion, Recte* [on-line]. 2005. [Dostęp 9 listopada 2007]. Dostępny w World Wide Web: <http://delos-wp5.ukoln.ac.uk/dissemination/pdfs/werla.pdf>.